

Virtual Hemispherical Amplitude Panning (VHAP): A Method for 3D Panning without Elevated Loudspeakers

Hyunkook Lee, Dale Johnson, and Maksims Mironovs

Applied Psychoacoustics Lab (APL), University of Huddersfield, Huddersfield, HD1 3DH, United Kingdom

Correspondence should be addressed to Hyunkook Lee (h.lee@hud.ac.uk)

ABSTRACT

This paper proposes ‘virtual hemispherical amplitude panning (VHAP)’, which is an efficient 3D panning method exploiting the phantom image elevation effect. Research found that a phantom centre image produced by two laterally placed loudspeakers would be localised above the listener. Based on this principle, VHAP attempts to position a phantom image over a virtual upper-hemisphere using just four ear-level loudspeakers placed at the listener’s left side, right side, front centre and back centre. A constant-power amplitude panning law is applied among the four loudspeakers. A listening test was conducted to evaluate the localisation performance of VHAP. Results indicate that the proposed method can enable one to locate a phantom image at various spherical coordinates in the upper hemisphere with some limitations in accuracy and resolution.

1 Introduction

Several past studies [1-6] reported that, when two identical signals are simultaneously reproduced from a pair of loudspeakers that are placed at ear level and arranged symmetrically from the listener position, the resulting phantom centre image would be perceived to be elevated in the median plane. It was also confirmed in the studies that the degree of perceived elevation would increase as the loudspeaker base angle increased from 0° to 180°; the image would be perceived almost right above the listener’s head when the base angle is 180°. Lee [5,6] recently found that the strength of this effect significantly depends on the type of sound source; sound sources with a flatter frequency spectrum and more transient nature would be perceived to be more elevated. There exists several theoretical explanations for this effect: head-rotation-dependent interaural time difference matching between phantom and real sources [1,2], the spectral energy distribution of the ear-input signals at high frequencies [4], matching between spectral notch cues of phantom and real sources at low frequencies [5,6].

Recent loudspeaker formats for immersive sound reproduction such as Dolby Atmos¹ and Auro-3D² employ the so-called height channels. Currently one of the most popular panning methods used for such systems is Vector Base Amplitude Panning (VBAP) [7]. However, in home environments it is often difficult to place loudspeakers

¹ www.dolby.com/atmos

² www.auro-3d.com

in elevated positions. The current study proposes a three-dimensional amplitude panning method that exploits the phantom image elevation effect, thus not requiring height channels. The aforementioned studies on the effect focused only on the elevation of a phantom centre image without applying any interchannel difference between loudspeaker signals. However, the method proposed in this paper, named 'Virtual Hemispherical Amplitude Panning (VHAP)', attempts to pan the elevated phantom image over the virtual upper hemisphere by utilising interchannel level differences (ICLDs) among four horizontally placed loudspeakers.

The current paper first presents the working principle of VHAP. It then describes a listening experiment conducted to evaluate the performance of VHAP.

2 Virtual Hemispherical Amplitude Panning (VHAP)

VHAP is a constant-power amplitude panning method that attempts to pan a phantom image over a virtual upper-hemisphere. It requires four ear-level loudspeakers placed at the listener's side left (SL), side right (SR), front centre (FC) and back centre (BC). The perceptual basis for this method is as follows. As a greater ICLD is applied between SL and SR with a constant power panning, the phantom image migrates from above to a fully left or right position in an arc trajectory rather than a straight line across the left and right. With a certain image position rendered in the lateral plane between SL and SR, a constant power weighting is applied between the SL-SR pair and FC for panning in the front half of the hemisphere. As the FC weighting is increased, the phantom image tends to be shifted from the initial position on the lateral plane towards FC, again in an arc trajectory. For panning in the back half of the upper hemisphere, simply BC is used instead of FC.

From this principle, a simple constant power panning law was derived for VHAP. The phantom image can be positioned arbitrarily on the virtual hemisphere by applying gain coefficients to the four loudspeaker signals for target spherical coordinates (azimuth and elevation). It should be noted that the main goal of VHAP is not to achieve a high accuracy in localisation, but to achieve 3D panning with a reasonable resolution, since the method does not attempt to physically reconstruct the exact ear-signal for a target source position, but rather relies on a psychoacoustic phenomenon.

Also it is worth noting that although three loudspeakers are involved for panning as in the 3D VBAP method [7], the proposed method differs from it in that (i) the loudspeakers are placed in the horizontal plane for elevation panning, (ii) the gain coefficients are not calculated from vectors oriented from the listener position, but from the spherical coordinate for the target image position on the virtual hemisphere within the loudspeaker layout, as described below.

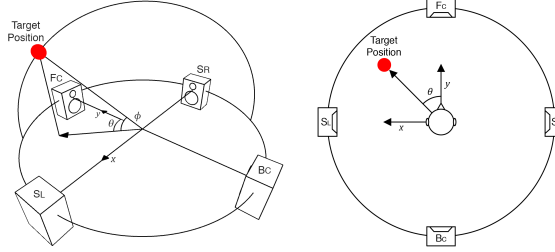


Figure 1. Aerial and top views of loudspeaker arrangement for VHAP.

Panning between all loudspeakers must satisfy the constant power panning law:

$$g_{SL}^2 + g_{SR}^2 + g_{FC}^2 + g_{BC}^2 = 1 \quad (1)$$

where g_{SL} , g_{SR} , g_{FC} , g_{BC} are gain coefficients for SL, SR, FC and BC, respectively.

Since VHAP uses a 2D circular speaker setup, the spherical coordinates must be converted and flattened down to a 2D co-ordinate that lies within the setup circle as can be seen in Figure 1.

$$x = \sin \theta \cdot \cos \phi \quad (2)$$

$$y = \cos \theta \cdot \cos \phi \quad (3)$$

where θ and ϕ are target azimuth and elevation in degrees, respectively, and x and y are the two components of the co-ordinate ranging from -1 to 1.

In VHAP, a maximum of three loudspeakers are in operation; either FC or BC is chosen depending on whether the target position lies within the front half or the back half of the virtual hemisphere. General equations for a constant power panning among three loudspeakers are given by

$$g_1 = \cos(a \cdot 90^\circ) \cdot \sin(b \cdot 90^\circ) \quad (4)$$

$$g_2 = \sin(a \cdot 90^\circ) \cdot \sin(b \cdot 90^\circ) \quad (5)$$

$$g_3 = \cos(b \cdot 90^\circ) \quad (6)$$

where a is a normalised weighting coefficient ranging between 0 and 1 for the balancing between SL and SR, while b is a normalised weighting coefficient between 0 and 1 for the SL-SR pair and FC or BC, depending on where the target position lies.

We now map the 2D coordinates x and y from (2) and (3) with the constant power gain coefficients from (4) to (6). The normalised weighting factor a can be defined as

$$a = \frac{x + x_{max}}{2x_{max}} \quad (7)$$

where x_{max} is the maximum possible value for x for a given y value, given by

$$x_{max} = \sqrt{1 - y^2} \quad (8)$$

x_{max} is also identical to the sine component of Equations (4) and (5).

$$x_{max} \equiv \sin(b \cdot 90^\circ) \quad (9)$$

The absolute value of y component is identical to g_3 from (6).

$$|y| \equiv \cos(b \cdot 90^\circ) \quad (10)$$

From the above, the gain coefficients for SL, SR, FC and BC are defined as follows.

$$g_{SL} = \sqrt{1 - y^2} \cdot \cos(a \cdot 90^\circ) \quad (11)$$

$$g_{SR} = \sqrt{1 - y^2} \cdot \sin(a \cdot 90^\circ) \quad (12)$$

$$g_{FC} = \begin{cases} |y|, & y \geq 0 \\ 0, & y < 0 \end{cases} \quad (13)$$

$$g_{BC} = \begin{cases} 0, & y \geq 0 \\ |y|, & y < 0 \end{cases} \quad (14)$$

3 Experiment

3.1 Method

A listening test was conducted in order to evaluate the performance of VHAP. It took place in an ITU-R BS.1116-compliant listening room at the Applied Psychoacoustics Lab (APL) of the University of Huddersfield. Four Genelec 8040A loudspeakers were arranged as in Figure 1. The distance of each loudspeaker from the listener position was 2m.

A total of 37 stimuli with different target spherical coordinates were created using a dry helicopter recording. The spherical coordinates tested were combinations of the azimuth angles from -150° to 180° with 30° intervals and the elevation angles of 0° , 30° , 60° and 90° .

10 subjects participated in the listening test, in which they tested each condition twice in a randomised order. They comprised staff and postgraduate researchers and undergraduate students from the University of Huddersfield's APL and music technology course. Their ages ranged from 21 to 40 and all reported to have normal hearing.

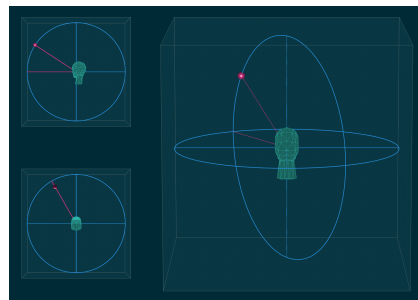


Figure 1. GUI used for the 3D localisation test. The azimuth and elevation of the dot in the 3D view are controlled using a rotation knob.

The subject's task was to localise the perceived position of each stimulus using a new graphical user interface (GUI) for 3D localisation test written in Matlab [8], which is shown in Figure 2. This GUI allows the user to locate a marker at a certain position on a sphere by individually controlling the azimuth and elevation of the marker using keyboard shortcuts or a rotation knob. The top-down and side views are also provided so that the subject can monitor where the exact azimuth and elevation of the marker are as they change them on the sphere.

The subjects repeated the localisation task twice for each stimulus in a randomised order. The test was split into four sessions, with each lasting an average time duration of 15 minutes. The subject's head movement was not allowed and this was strictly monitored using a head-tracking system; if the subject's head moved more than 1° , the audio stopped automatically and the GUI showed an instruction to help the subjects move back to the correct head position, after which the audio playback could resume.

3.2 Results

Shapiro-Wilks test suggested that the data collected were largely non-normally distributed, and therefore non-parametric methods were used for statistical analysis. The data were also examined for bimodality using Hartigan's dip test. It was found that

a number of azimuth conditions had a significant bimodality ($p < 0.05$), exhibiting a front-back confusion tendency. No elevation condition had a significant bimodality.

Table 1. Median perceived azimuth and elevation for each target azimuth (TA) and target elevation (TE).

TA	TE			
	0°	30°	60°	90°
0°	0°, 1°	0°, 12°	BM, 49°	BM, 56°
30°	26°, 20°	26°, 33°	38°, 52°	-
60°	60°, 23°	48°, 29°	BM, 42°	-
90°	90°, 0°	90°, 1°	BM, 47°	-
120°	123°, 23°	BM, 27°	BM, 48°	-
150°	155°, 30°	BM, 38°	BM, 50°	-
180°	180°, 7.5°	BM, 37°	BM, 56°	-

The data obtained for each target position are plotted in Figure 3. For the conditions with non-bimodal azimuth perception, the data are presented as scatter plots as well as two-dimensional notch error bars. The notch range is a non-parametric equivalent of 95% confidence interval. The point where the two bars cross indicate median perceived azimuth (PA) and perceived elevation (PE). The solid lines show the target azimuth (TA) and target elevation (TE). In general, if the notch error bar overlaps with the target lines, it can be considered that the difference is non-significant [9]. For the conditions with bimodality, the median and notch error bar of elevation data are plotted separately from the scatter plots.

For the conditions with the TA of 0°, it appears from Figure 3 that the PA was accurate for the TEs of 0° and 30°, but the TEs of 60° and 90° produced a significant front to back confusion. The PE for the 0° TA increased as the TE increased, although the PE was lower than the target; the range of median PE was from 1° to 56° for the TE from 0° to 90°.

For the 30° TA conditions, both PA and PE are close to the target in general without any bimodality, although the median PE was 20° when TE was 0°.

For the 60° TA, the 0° TE condition was elevated to the median PE of 23° similar to the TA30°/TE0° condition. The 30° TE condition was perceived around the target. The 60° TE condition had a bimodality in PA and its median PEs was 42°.

The 90° TA conditions did not elevate the perceived image for the TEs of 0° and 30°, but the 60° TE was elevated to the median of 50°, although the PA was bimodal between 30° and 150°.

For all of the TAs in the back half (120° , 150° and 180°), all of the TEs apart from 0° produced a bimodal distribution in PA. The azimuth of the 0° TE condition was mostly perceived correctly, but the median PEs for the 120° and 150° TEs were 23° and 30° , respectively. For all of these TAs, the median PEs for the 30° and 60° TEs were reasonably close to the target.

3.3 Discussion

The results generally suggest that VHAP is able to produce the virtual elevation of a phantom image at various spherical coordinates with some limitations in accuracy and resolution. For the 30° and 60° TEs in particular, the PEs generally had a good match with the TEs for most of the TAs.

However, with respect to PA, the number of conditions with bimodality increased as the TA and TE increased. For the 30° TE, a significant back-to-front confusion was observed only for the TAs greater than 90° , but for the 60° and 90° all conditions apart from the 30° TA had a significant front-to-back or back-to-front confusion. This seems to suggest an elevation dependency of the front-back confusion phenomenon in general.

It is also noticeable from the results that the conditions with the 0° TE still had the PE around 20° - 25° , except for the ones that had hard panning to a real source (e.g., the TAs of 0° , 90° and 180°). This indicates that the phantom image elevation effect still operates with a loudspeaker pair that is asymmetrical from the listener position.

Although the method described here involves a specific loudspeaker arrangement, it is expected that VHAP would still be found useful for the conventional 5.1 or 7.1 setup. For instance, when there is no back centre loudspeaker available, using VHAP for the front centre and side or rear channels of a 5.1 or 7.1 layout would allow one to create elevated images in the front half of the virtual hemisphere. This would be useful for the downmixing of 3D audio content, especially in an object-based audio scenario, as well as spatial sound design and music composition without height channels.

The experiment described in this study only tested static target positions. The localisation of a phantom source in a 3D space is inherently a difficult task, which is a potential reason for the rather large data spread. Furthermore, there is a potential influence of the spectral characteristics of the source used on the results (e.g. pitch-height effect). However, for a practical use scenario, the user would use a GUI to control over the position of the panned image, regardless of what the original target position is. For this purpose, a VHAP GUI written in Max is freely available for download from the resources section at www.hud.ac.uk/apl.

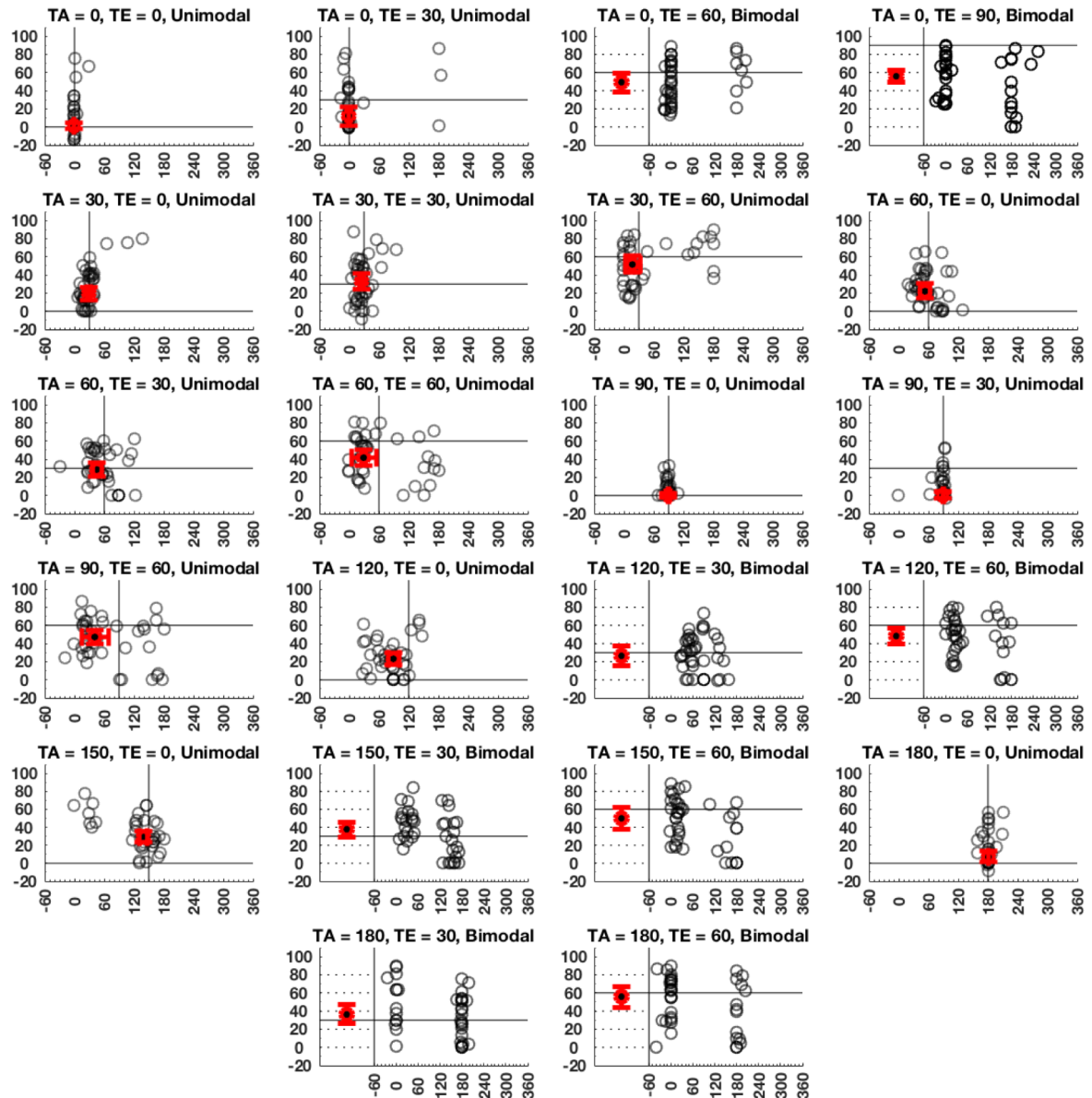


Figure 3. Plots of perceived azimuth and elevation from the listening test. TA = Target Azimuth and TE = Target Elevation in degrees. For conditions with unimodal azimuth perception, median and two-dimensional notch ranges are plotted over the scatterplots. For conditions with bimodal azimuth perception, the median and notch range for perceived elevation is plotted separately from the scatter plots.

4 Conclusion

This paper proposes a new amplitude panning method for creating virtually elevated image over the upper-hemisphere, named ‘virtual hemisphere amplitude panning (VHAP)’. It applies constant power panning among three loudspeakers: side left, side right and front centre or back centre depending on whether the target coordinate is in the front or back. The results from the listening test indicate that VHAP achieves the goal with limitations in accuracy and resolution. Future works will include the testing of other sound sources.

References

- [1] K. de Boer, “A Remarkable Phenomenon with Stereophonic Sound Reproduction,” *Philips Tech. Rev.*, vol. 9, pp. 8–13 (1947).
- [2] D. M. Leakey, “Some Measurements on the Effects of Interchannel Intensity and Time Differences in the Two Channel Sound Systems,” *J. Acoust. Soc. Am.*, vol. 31, pp. 977–986 (1959). doi.org/10.1121/1.1907824
- [3] P. Damaske and V. Mellert, “A Procedure for Generating Directionally Accurate Sound Images in the Upper Half-Space Using Two Loudspeakers,” *Acustica*, vol. 22, pp. 154–162 (1969/1970).
- [4] J. Blauert, *Spatial Hearing*, rev. ed. (MIT Press, Cambridge, MA, 1997).
- [5] H. Lee, “Sound Source and Loudspeaker Base Angle Dependency of Phantom Image Elevation Effect” *J. Audio Eng. Soc.*, vol. 65, pp. 733–748 (2015 Jul./Aug.) doi.org/10.17743/jaes.2017.0028
- [6] H. Lee, “Phantom Image Elevation Explained,” presented at the *141st Convention of the Audio Engineering Society* (2016 Sep.), convention paper 9664.
- [7] V. Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning,” *J. Audio Eng. Soc.*, vol. 45, pp. 456–466 (1997 Jun.).
- [8] M. Mironovs and H. Lee, “On the accuracy and consistency of sound localisation at various azimuth and elevation angles”, presented at the *144th Convention of the Audio Engineering Society* (2018 May).
- [9] R. McGill, J. W. Tukey, W. A. Larsen, “Variations of Box Plots”. *Am. Stat.*, vol. 32, pp. 12–16 (1978).